The Road to DocuComp

Ted Habermann, NOAA/NESDIS/NGDC

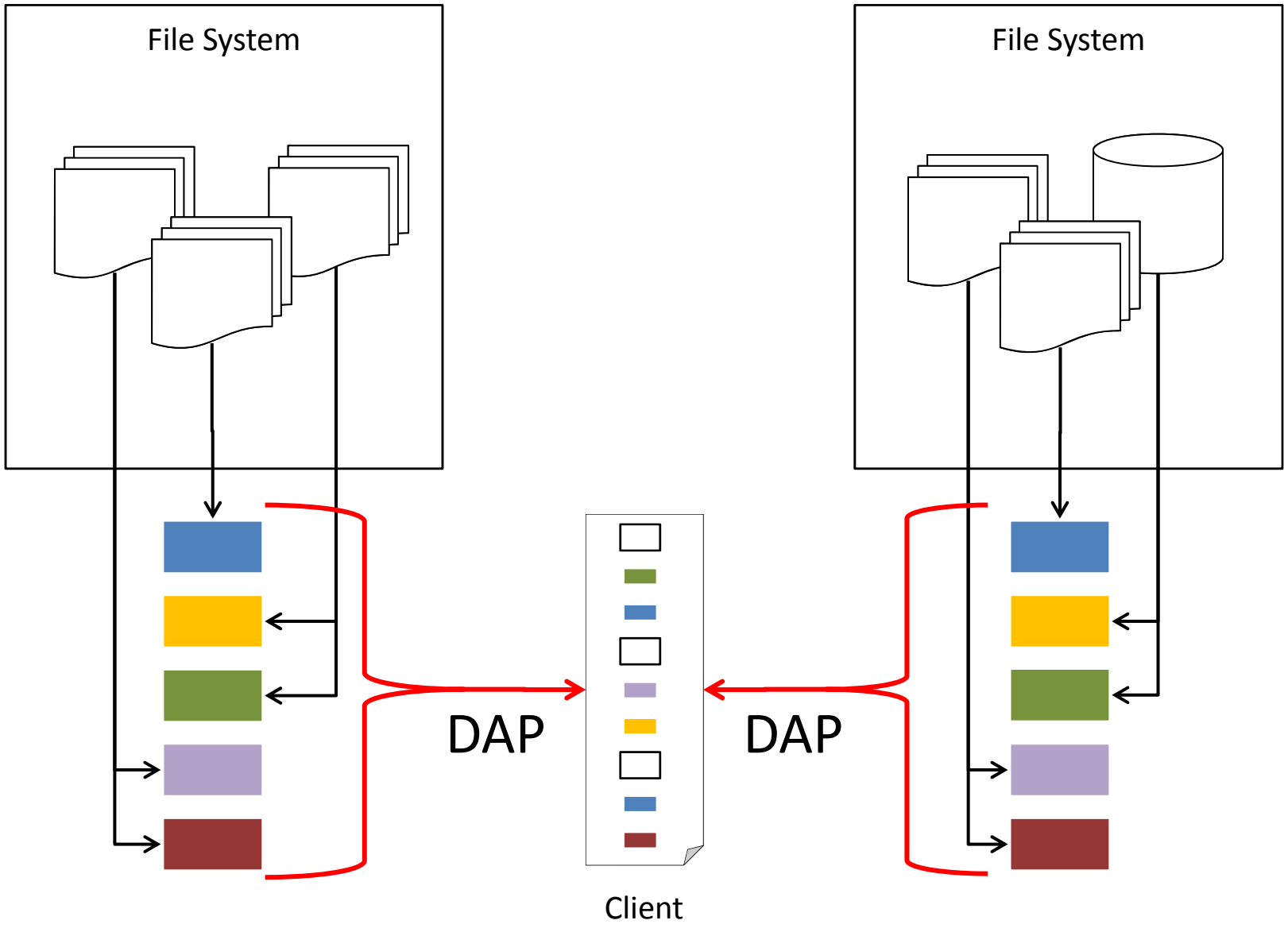
NCDC Metadata Workshop

Database normalization is the process of organizing the fields and tables of a relational database to minimize redundancy and dependency.

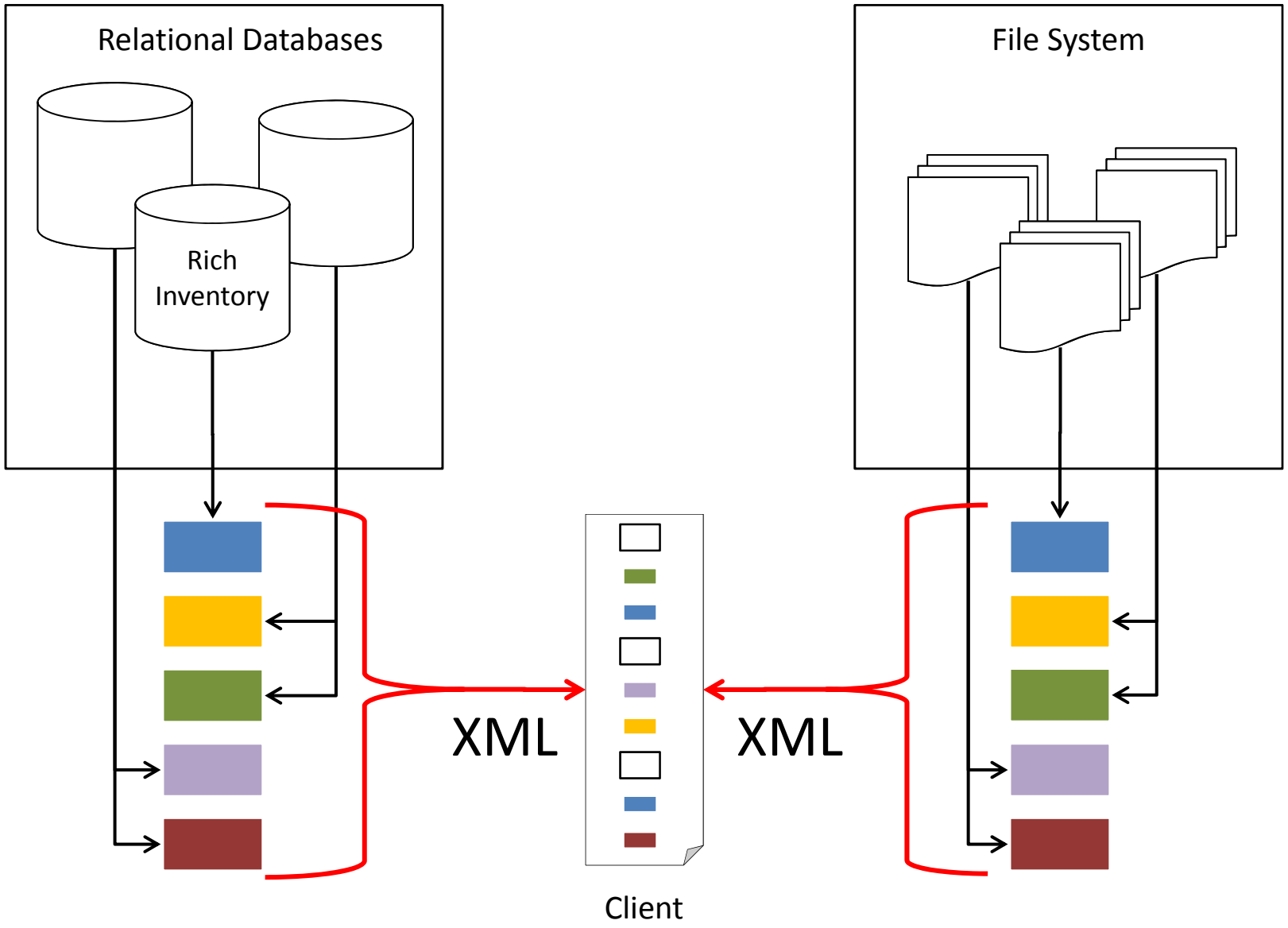
The objective is to isolate data so that additions, deletions, and modifications can be made once and then propagated through the rest of the database via defined relationships.

DocuComp brings the benefits of normalization to metadata using reusable components and xlinks instead of tables and foreign keys.

Persistence vs. Transport – OPeNDAP



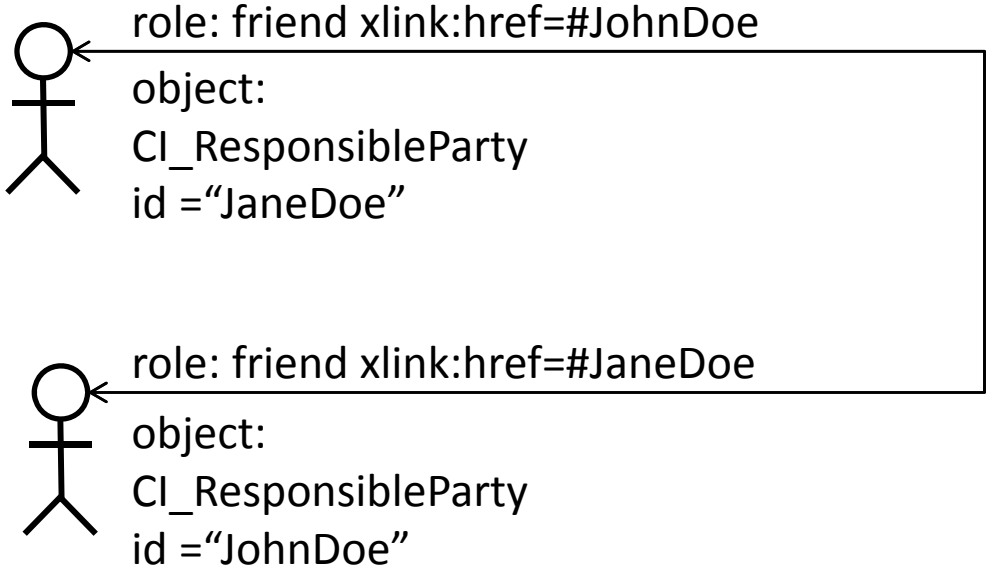
Persistence vs. Transport - Documentation



Roles, Identifiers and References

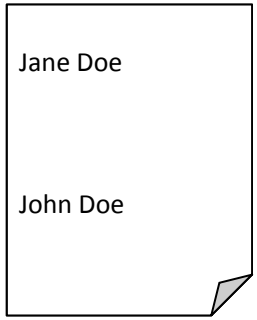
ID	Name	Friend_ID
JaneDoe	Jane Doe	JohnDoe
MaryBrown	Mary Brown	TomBrown
TomBrown	Tom Brown	MaryBrown
JohnDoe	John Doe	JaneDoe

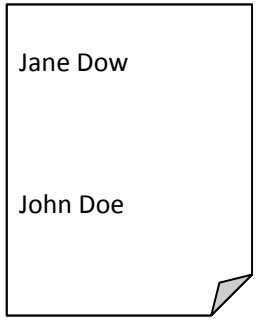
This table has four people, all of which have identifiers (ID), names, and friends. The friends are identified by foreign keys (Friend_ID).

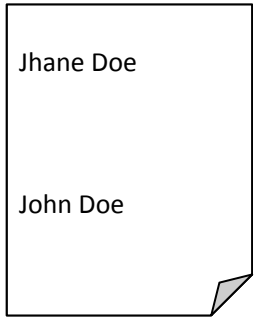


This information is represented in XML as objects with ids and references with xlinks.

Jane or Jhane?





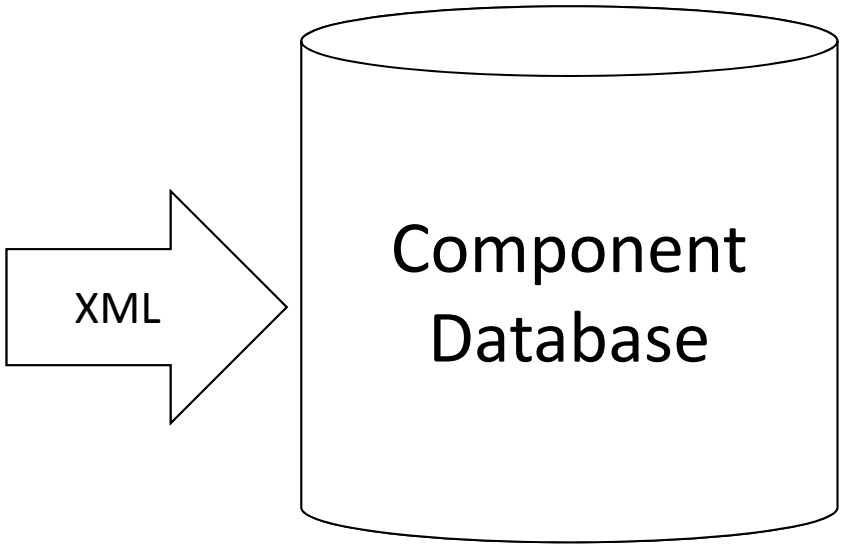



```
<gmd:contact>
 <gmd:CI_ResponsibleParty>
  <gmd:individualName>
   <gco:CharacterString>Jhane Doe</gco:CharacterString>
  </gmd:individualName>
  <gmd:organisationName>
   <gco:CharacterString>
   DOC/NOAA/NESDIS/NGDC > National Geophysical Data Center, NESDIS, NOAA, U.S. Department of Commerce
   </gco:CharacterString>
  </gmd:organisationName>
  <gmd:contactInfo>
   <gmd:CI_Contact>
    <gmd:address>
     <gmd:CI_Address>
      <gmd:electronicMailAddress>
       <gco:CharacterString>ngdc.info@noaa.gov</gco:CharacterString>
      </gmd:electronicMailAddress>
     </gmd:CI_Address>
    </gmd:address>
    <gmd:onlineResource>
     <gmd:CI_OnlineResource>
      <gmd:linkage>
       <gmd:URL>http://www.ngdc.noaa.gov</gmd:URL>
      </gmd:linkage>
     </gmd:CI_OnlineResource>
    </gmd:onlineResource>
   </gmd:CI_Contact>
  </gmd:contactInfo>
  <gmd:role>
   <gmd:CI_RoleCode codeListValue="pointOfContact">pointOfContact</gmd:CI_RoleCode>
  </gmd:role>
 </gmd:CI_ResponsibleParty>
</gmd:contact>
```

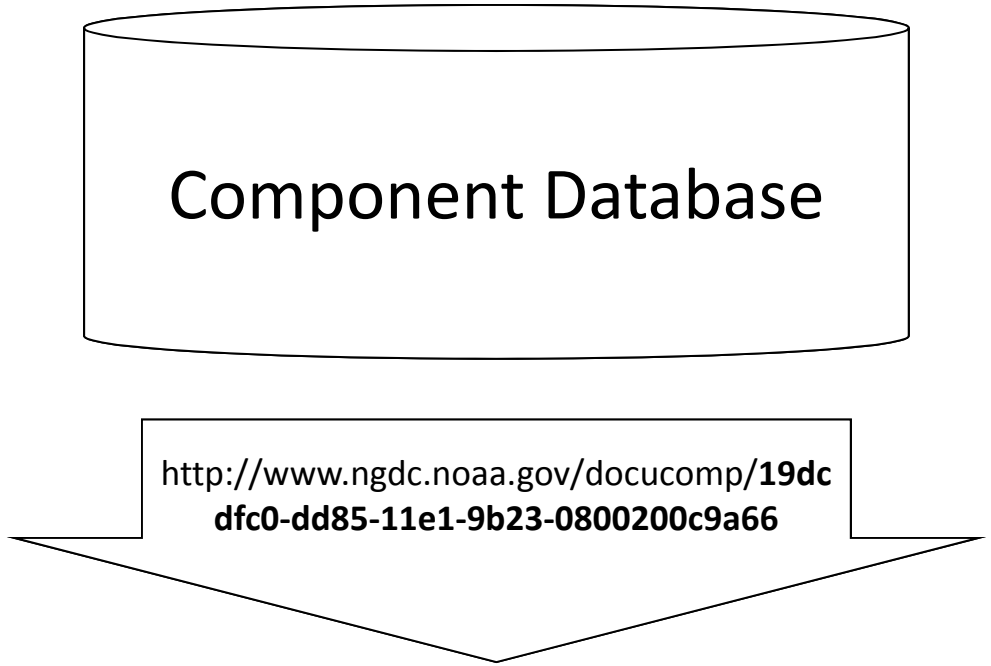
Jane as Reusable Component

```
<gmd:contact>
 <gmd:CI_ResponsibleParty UUID="19dcdfc0-dd85-11e1-9b23-0800200c9a66">
  <gmd:individualName>
   <gco:CharacterString>Jane Doe</gco:CharacterString>
  </gmd:individualName>
  <gmd:organisationName>
   <gco:CharacterString>
   DOC/NOAA/NESDIS/NGDC > National Geophysical Data Center, NESDIS, NOAA, U.S. Department of Commerce
   </gco:CharacterString>
  </gmd:organisationName>
  <gmd:contactInfo>
   <gmd:CI_Contact>
    <gmd:address>
     <gmd:CI_Address>
      <gmd:electronicMailAddress>
       <gco:CharacterString>ngdc.info@noaa.gov</gco:CharacterString>
      </gmd:electronicMailAddress>
     </gmd:CI_Address>
    </gmd:address>
    <gmd:onlineResource>
     <gmd:CI_OnlineResource>
      <gmd:linkage>
       <gmd:URL>http://www.ngdc.noaa.gov/mgg/</gmd:URL>
      </gmd:linkage>
     </gmd:CI_OnlineResource>
    </gmd:onlineResource>
   </gmd:CI_Contact>
  </gmd:contactInfo>
  <gmd:role>
   <gmd:CI_RoleCode codeListValue="pointOfContact">pointOfContact</gmd:CI_RoleCode>
  </gmd:role>
 </gmd:CI_ResponsibleParty>
</gmd:contact>
```

object:
CI_ResponsibleParty
id = Universally Unique ID (UUID)

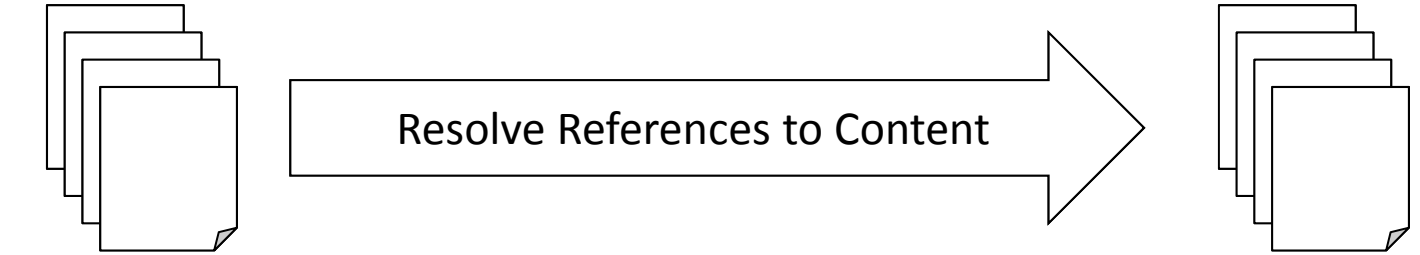


Docucomp = Database + Restful Service




```
<gmd:CI_ResponsibleParty UUID="19dcdfc0-dd85-11e1-9b23-0800200c9a66">
  <gmd:individualName>
    <gco:CharacterString>Jane Doe</gco:CharacterString>
  </gmd:individualName>
    …
</gmd:CI_ResponsibleParty>
```

Component Resolution

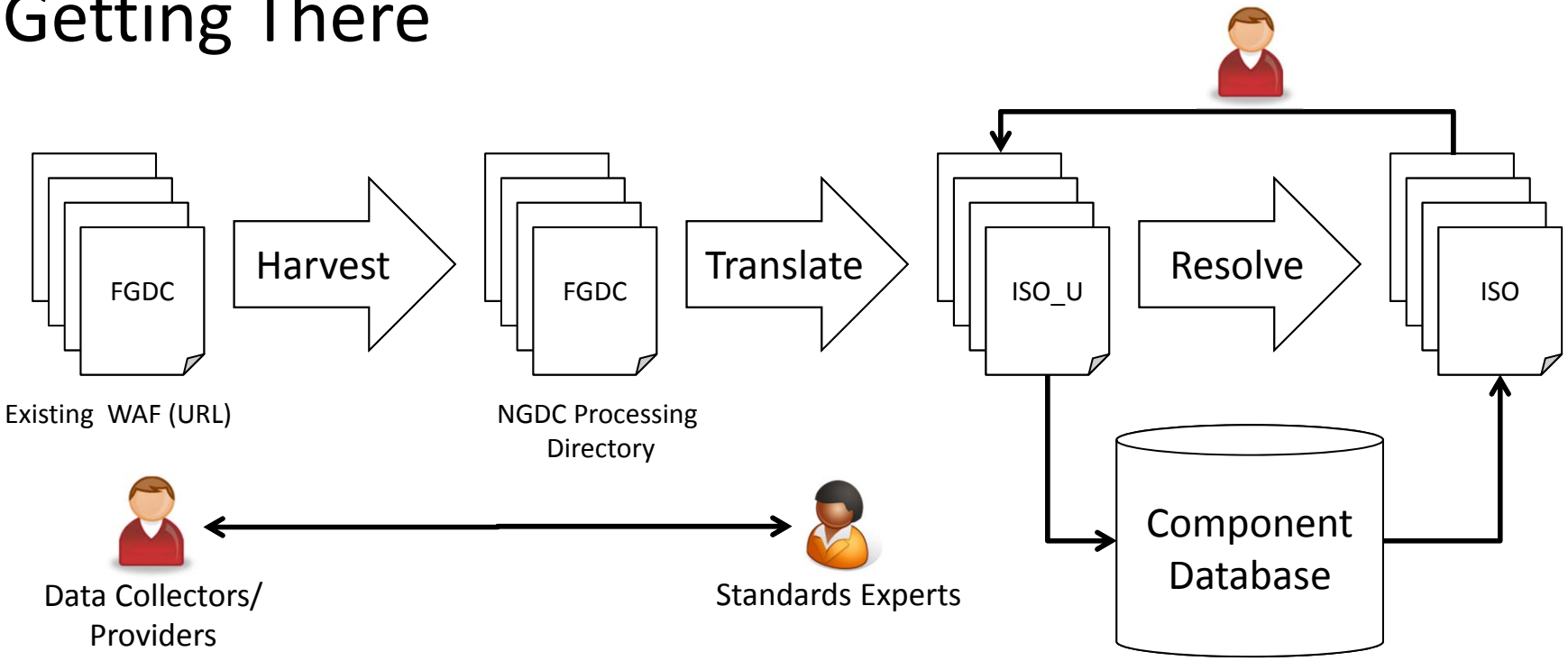


RecordSet/**iso_u**/xml

RecordSet/**iso**/xml

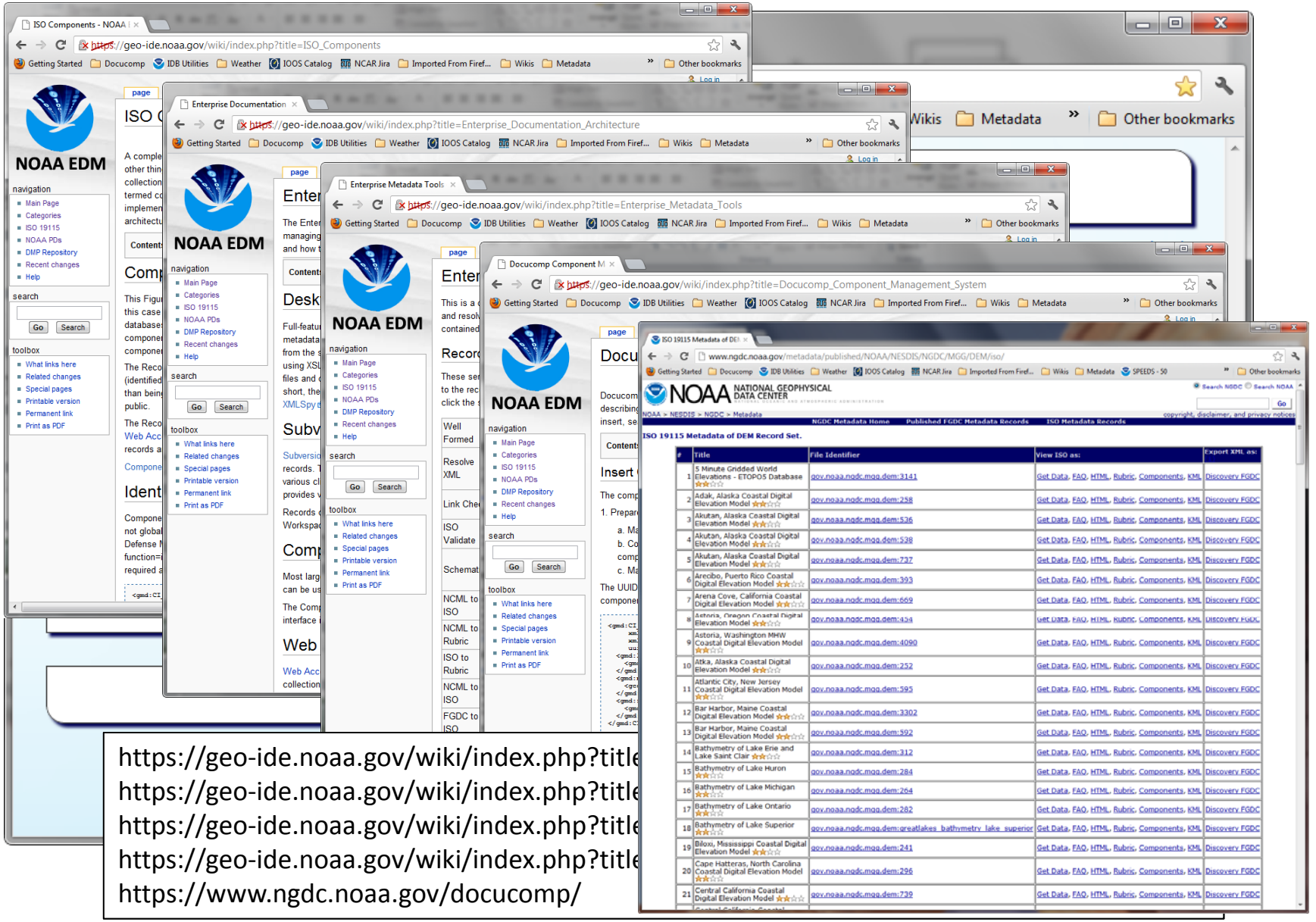
```
<gmd:contact xlink:href="http://www.ngdc.noaa.gov/docucomp/19dcdfc0-dd85-11e1-9b23-0800200c9a66" xlink:title="Jane Doe"/>
 <gmd:CI_ResponsibleParty UUID="19dcdfc0-dd85-11e1-9b23-0800200c9a66">
  <gmd:individualName>
    <gco:CharacterString>Jane Doe </gco:CharacterString>
  </gmd:individualName>
  <gmd:organisationName>
   <gco:CharacterString>
   DOC/NOAA/NESDIS/NGDC > National Geophysical Data Center, NESDIS, NOAA, U.S. Department of Commerce
   </gco:CharacterString>
  </gmd:organisationName>
  <gmd:contactInfo>
    …
 </gmd:contactInfo>
  <gmd:role>
   <gmd:CI_RoleCode codeListValue="pointOfContact">pointOfContact</gmd:CI_RoleCode>
  </gmd:role>
 </gmd:CI_ResponsibleParty>
</gmd:contact>
```

Getting There



Iterations:
Translation Proofing: identify un-translated content and fix
Rubrics: evaluate completeness and improvement opportunities
Validation: schematron and schema validation
Link checking: check all online resources
Multiple views: Translation of ISO to multiple views
Consistency Checker: find repeated content and candidate components
DocuComp: create and store components
Metadata editor: add references into metadata records

Resources


https://geo-ide.noaa.gov/wiki/index.php?title
https://geo-ide.noaa.gov/wiki/index.php?title
https://geo-ide.noaa.gov/wiki/index.php?title
https://geo-ide.noaa.gov/wiki/index.php?title
https://www.ngdc.noaa.gov/docucomp/

Questions?